A large-scale comparison of human-written versus ChatGPT-generated essays

Year: 2023 | Citations: 252 | Authors: Steffen Herbold, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva, Alexander Trautsch

Abstract

ChatGPT and similar generative AI models have attracted hundreds of millions of users and have become part of the public discourse. Many believe that such models will disrupt society and lead to significant changes in the education system and information generation. So far, this belief is based on either colloquial evidence or benchmarks from the owners of the models—both lack scientific rigor. We systematically assess the quality of AI-generated content through a large-scale study comparing human-written versus ChatGPT-generated argumentative student essays. We use essays that were rated by a large number of human experts (teachers). We augment the analysis by considering a set of linguistic characteristics of the generated essays. Our results demonstrate that ChatGPT generates essays that are rated higher regarding quality than human-written essays. The writing style of the AI models exhibits linguistic characteristics that are different from those of the human-written essays. Since the technology is readily available, we believe that educators must act immediately. We must re-invent homework and develop teaching concepts that utilize these AI models in the same way as math utilizes the calculator: teach the general concepts first and then use AI tools to free up time for other learning objectives.